IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

Multimodal Visual Search using CBIR system on Mobile Device

Prasad Kamthe[1], Aniruddha Bidnur[2], Rahul Patil[3], Sharad Pawar[4]

Information Technology Department, PVPIT Bavdhan, Pune, India[1,2,3,4]

Abstract: In this paper, it describe the content based image retrieval system based on three major types of visual information color, texture and shape, and their distances to the origin in a three dimensional space for the retrieval. It also describes a multimodal interactive image search system using mobile devices. This system useful for users who already have pictures in their minds but have no precise descriptions or names to address them. By describing it using speech and then refining the recognized query by interactively composing a visual query using exemplary images, the user can easily find the desired images through a few natural multimodal interactions with their mobile device.

Keywords: Multimodal visual search, interactive search, mobile device, Image Retrieval, Color Histogram.

I. INTRODUCTION

Research on content-based image retrieval has gained tremendous during the last decade. The term Content Based Image Retrieval (CBIR) for the automatic retrieval of the images from a database, based on the color and shape present. Since then, the term has widely been used to describe the process of retrieving desired images from a large collection of database, on the basis of syntactical image features (color, texture and shape).In many areas such as medicine, military, crime prevention, architecture, art and academic, large collections of digital images are being created. Many of these collections are the product of digitizing existing collections of photographs, diagrams, drawings, paintings, and prints. To access appropriate information, we need to retrieve these images from large image databases. Thus, image retrieval becomes an important issue. Image retrieval could be based on textual metadata or image content information.

Existing search alternatives for mobile users include text-based search and local map search. The user can either type an entity name or look up on a online local map to find the target. Moreover, photo-to-search is becoming pervasive as the development of the computer vision and content-based image retrieval. In this paper we enable the user to search the image using content based image retrieval system.

It is also observed that the users' search interest on the mobile device also differs from that on desktop. In CBIR, each image that is stored in the database has its features extracted and compared to the features of the query image. It involves two steps the first step in the process is extracting image features to a distinguishable extents. The second step involves matching these features to yield a result that is visually similar.

The basic fundamentals of content based image retrieval are divided into three parts feature extraction; multidimensional indexing and retrieval systems design. They discussed about the use of digital images and rapid increase in the size of digital image collections. The proper organization of the generated large amount of images by both military and civilian is needed, so as efficient browsing, searching and retrieval takes place. The two ways of image retrieval text based image retrieval and visual based image retrieval.

Image databases and collections can be enormous in size, containing hundreds, thousands or even millions of images. The conventional method of image retrieval is searching for a keyword that would match the descriptive keyword assigned to the image by a human categorizer. Currently under development, even though several systems exist, is the retrieval of images based on their content, called Content Based Image Retrieval, CBIR. While computationally expensive, the results are far more accurate than conventional image indexing. Hence, there exist tradeoffs between accuracy and computational cost. This tradeoffs decreases as more efficient algorithms are utilized and increased computational power becomes inexpensive.

This paper focuses on a low-dimensional shape based indexing technique for achieving efficient and effective retrieval performance. We present a simple index based on shape features of regions that are segmented out of images based on color. A new shape similarity measure conforming to human perception is applied and shown to be effective. Images are segmented to obtain homogeneous color regions that are dominant and similar images form an image cluster stored in a hash structure. Each region within an image is then indexed by a region-based shape index. The shape index is invariant to translation, rotation and scaling. A JAVA based query engine supporting query-by-example is developed for retrieving images by shape. The retrieval performance is studied and compared with that of a region-based shape-indexing scheme.

The whole search interaction begins with the natural language understanding. In this paper, a simple entity extraction strategy is adopted to handle this problem.The entity extraction from speech can be divided into two steps:

i. Speech recognition
ii. Key entity extraction.

Copyright to IJARCCE

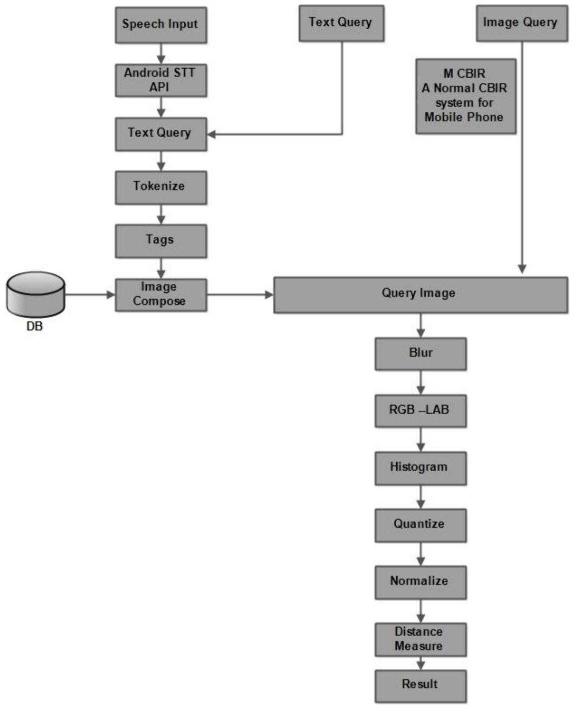

Fig 1.1 Architecture

After the voice is captured by the phone, it is first sent to commercial online speech recognition engine to translate the recorded audio stream into a text string. A Hidden Markov Model (HMM) and N-gram-based engine is able to handle both natural sentences and phrase fragments. However, it is preferred to use natural sentences like “find an image with iron tower on the grass” instead of separate phrases like “iron tower, grass”. Although a common text-based image retrieval engine can perform better with unnatural phrase segments as “iron tower” and “grass,” a whole sentence is easier tube recognized by speech recognizer and much more natural for the user. Actually, an entire sentence is also easier to be recognized by speech engine because the speech engine uses Markov-based algorithm to recognize the words within audio stream, a complete sentence can be recognized with higher accuracy. For example, it’s hard to decide whether to choose “abbey” or “obey” if only a single word is uttered. But within sentence, the language model can choose the suitable words according to the context and part of speech.

II. LAB COLOR MODEL

A *Lab* color space is a color-opponent space with dimension *L* for lightness and *a* and *b* for the color-opponent dimensions, based on nonlinearly compressed CIE XYZ color space coordinates. The dimensions of the Hunter 1948 *L*, *a*, *b* color space are *L*, *a*, and *b*.[1][2] However, *Lab* is now more often used as an informal abbreviation for the CIE 1976 (*L**, *a**, *b**) color space (or CIELAB). The difference between Hunter and CIE color coordinates is that the CIE coordinates are based on a cube root transformation of the color data, while the Hunter coordinates are based on a square root transformation. The L*a*b* color space includes all perceivable colors, which means that its gamut exceeds those of the RGB and CMYK color models (for example, ProPhoto RGB includes about 90% all perceivable colors). One of the most important attributes of the L*a*b*-model is device independence. This means that the colors are defined independent of their nature of creation or the device they are displayed on. The L*a*b* color space is used when graphics for print have to be converted from RGB to CMYK, as the L*a*b* gamut includes both the RGB and CMYK gamut. Also it is used as an interchange format between different devices as for its device independency.

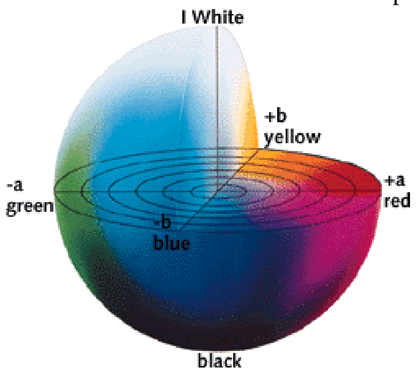

Fig 1.2 LAB color model

III. CONCLUSION

We have introduced an interactive mobile visual search system which allows the users to formulate their search intent through natural multimodal interactions with mobile devices. The system first represents the study on mobile visual search by taking the advantages of multimodal and multitouch functionalities on the phone. The visual query generated by the user can be effectively used to retrieve similar images. We proposed combining color, shape, texture features for content based image retrieval. We have experimentally investigated several feature extraction methods with several learning algorithm.

ACKNOWLEDGEMENT

We take this opportunity to thank our project guide **Prof. P. P. Waghralkar** and Head of the Department **Prof. N. D. Kale** for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this seminar report. We are also thankful to **Prof. S. B. Madankar** and all the staff members of the Department of Information Technology of PVPIT, Bavdhan Pune. For their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

REFERENCES

[1] Houqiang Li, Yang Wang, Tao Mei ,Jingdong Wang , and Shipeng Li , ” Interactive Multimodal Visual Search on Mobile Device”, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 3, APRIL 2013.

[2] Pratheep Anantharatnasamy, Kaavya Sriskandaraja, Vahissan Nandakumar and Sampath Deegalla,” Fusion of Color, Shape and Texture Features for Content Based Image Retrieval”, The 8th International Conference on computer Science & Education (ICCSE 2013) April 26-28, 2013. Colombo, Sri Lanka.

[3] Neetu Sharma, Paresh Rawat and jaikaran Singh.S3,” Efficient CBIR Using Color Histogram processing”, “Signal & Image Processing: An International Journal (SIPIJ) Vol.2, No.1, March 2011”.

[4] E. Rublee, V. Rabaud, K. Konolige, and G.Bradski,“ORB: An efficient alternative to SIFT or SURF,” in Proc. IEEE Int. Conf. Compute. Vis., Nov. 2011, pp. 2564–2571.

[5] Y.Wang, T.Mei, J.Wang, H. Li, and S. Li, “JIGSAW: Interactive mobile visual search with multimodal queries,” in *Proc. ACMMultimedia*, 2011, pp. 73–82.